MAGI as a Geometric Generalization of Momentum-Based Optimization

Flyxion

November 27, 2025

1 Introduction

Classical optimization methods in machine learning treat parameter spaces as undifferentiated Euclidean domains. Gradient descent and its momentum-based variants, including Stochastic Gradient Descent with Momentum (SGDM) Polyak1964,Sutskever2013, assume that all directions in the ambient space are semantically valid and that the geometry of meaningful representation plays no explicit role in the learning process. The resulting algorithms are powerful but exhibit characteristic failure modes: drift into regions with no coherent interpretation, oscillation in directions orthogonal to underlying data manifolds, and sensitivity to local distortions of coordinate systems.

The Manifold-Aligned Generative Inference (MAGI) framework provides a structured alternative. MAGI is built upon the premise that learning does not occur in an arbitrary Euclidean space but instead unfolds along a stratified Riemannian semantic manifold that captures the intrinsic geometry of meaningful representations. This manifold, denoted $M \subset \mathbb{R}^n$, is Whitney-stratified Whitney1965,Thom1969,Mather1973 and equipped with a family of Riemannian metrics across its strata LeeRiemannian2018,doCarmoRiemannian1992. Learning proceeds through a constrained form of heavy-ball dynamics Polyak1964,Wibisono2016 that projects gradients onto tangent spaces, suppressing motion orthogonal to the manifold and ensuring that updates remain coherent with the underlying semantic structure.

A central contribution of this chapter is the demonstration that SGDM arises naturally as a degenerate case of MAGI when all geometric structure is suppressed. When M is taken to be the entire ambient space, when the stratification is trivial GoreskyMacPherson1988, and when the tangent projection operator is the identity, the MAGI update reduces exactly to the SGDM update. This containment relation reveals SGDM as a boundary case of a more general geometric theory.

The geometric enhancements introduced by MAGI provide stability properties that SGDM cannot match. In particular, the explicit suppression of normal components—made precise

using classical submanifold geometry doCarmoRiemannian1992,CheegerEbin1975—ensures that motion orthogonal to the semantic manifold is eliminated, whereas SGDM permits such components to accumulate through the momentum term. A formal theorem presented later in the chapter proves that the normal component of the gradient cannot increase under the MAGI update except through the intrinsic variation of the potential along tangential directions.

The remainder of this chapter develops the theory in detail. The next section provides differential geometric preliminaries, including Riemannian manifolds LeeSmooth2013,LeeRiemannian2018, tangent and normal bundles doCarmoRiemannian1992, exponential maps CheegerEbin1975, and curvature effects Sakai1996. Subsequent sections introduce Whitney stratification Whitney1957,Whitney1965, stratified Morse potentials Milnor1963,Goresky-MacPherson1988, and their relevance to structured learning dynamics. The Riemannian heavy-ball equation is derived and compared to its Euclidean counterpart Polyak1964, establishing the analytic foundation for the MAGI update rule. The MAGI dynamics are then developed formally, followed by proofs of their structural properties and their relationship to SGDM. The chapter concludes with an extended discussion of the implications of this framework for geometry-aware learning and manifold-constrained optimization Absil2008,Boumal2023.

2 Differential Geometric Preliminaries

The MAGI framework rests on the geometric structure of smooth manifolds, Riemannian metrics, and submanifold theory. The material in this section follows standard references in differential geometry LeeSmooth2013,LeeRiemannian2018,doCarmoRiemannian1992,Jost2011,Sakai1996. We summarize the concepts required for the construction of the MAGI update rule while emphasizing those geometric properties that directly influence optimization dynamics.

Smooth Manifolds and Riemannian Structure

A smooth manifold S of dimension d is a Hausdorff space locally diffeomorphic to $\mathbb{R}^d$; the charts and transition maps endow S with its differentiable structure LeeSmooth2013. A Riemannian metric g assigns to each point $x \in S$ an inner product g_x on the tangent space $T_x S$, varying smoothly in x. The metric induces norms, angles, lengths of curves, and provides the setting for geodesic motion LeeRiemannian2018.

The Levi-Civita connection ∇ is the unique torsion-free connection compatible with the metric, satisfying

$$X\langle Y, Z \rangle = \langle \nabla_X Y, Z \rangle + \langle Y, \nabla_X Z \rangle.$$

This connection governs the behavior of vector fields along curves and determines the

geodesics of (S, g).

Exponential Map and Geodesic Flow

For each $x \in S$, the exponential map $_x : T_x S \to S$ is defined by

$$_x(v) = \gamma_v(1),$$

where γ_v is the unique geodesic satisfying $\gamma_v(0) = x$ and $\gamma_v'(0) = v$. Existence and smoothness of the exponential map follow from standard results in Riemannian geometry CheegerEbin1975,LeeRiemannian2018. In Euclidean space, $_x(v) = x + v$, while on curved manifolds, the exponential incorporates curvature-dependent distortions.

Classical comparison theorems CheegerEbin1975,Sakai1996 ensure that for sufficiently small $\|v\|$, the exponential map is a diffeomorphism onto a normal neighborhood of x, enabling local coordinate descriptions of geodesics and gradient flows.

Submanifolds, Tangent Spaces, and Normal Spaces

Let $M \subset \mathbb{R}^n$ be a smoothly embedded submanifold of dimension d. By the submanifold theorem doCarmoRiemannian1992,GuilleminPollack1974, for each $x \in M$ there exists an open neighborhood diffeomorphic to a product of tangent and normal components. The tangent space $T_x M$ is the set of velocities of curves in M passing through x, while the normal space is defined by

$$N_x M = (T_x M)^\perp$$

with respect to the ambient Euclidean inner product.

Every vector $u \in \mathbb{R}^n$ admits a unique orthogonal decomposition

$$u = \Pi_{T_x M}(u) + \Pi_{N_x M}(u),$$

where $\Pi_{T_x M}$ and $\Pi_{N_x M}$ are orthogonal projections. These projections depend smoothly on x when M is smooth.

The second fundamental form $II_x : T_x M \times T_x M \to N_x M$ encodes the curvature of M inside the ambient space and governs the rate at which normal components emerge along tangent flows doCarmoRiemannian1992. This curvature dependence plays a role later in bounding the intrinsic normal drift under MAGI.

Retractions and Local Approximations

Although the exponential map provides the canonically correct update rule in Riemannian optimization, it is often expensive to compute. A smooth mapping $R_x : T_x S \to S$ is called

a retraction if

$$R_x(0) = x, \qquad DR_x(0) = \mathrm{id}_{T_x S},$$

and it approximates the exponential map to first order Absil2008,Boumal2023. Retractions offer computationally efficient substitutes for Riemannian geodesic updates. For MAGI, however, we retain the exact exponential map because the geometric semantics motivate strict consistency with manifold structure.

Gradient Fields on Riemannian Manifolds

For a smooth function $F : S \to \mathbb{R}$, the Riemannian gradient $\nabla F(x)$ is defined implicitly by

$$\langle \nabla F(x), v \rangle_x = dF_x(v) \qquad \text{for all } v \in T_x S.$$

Gradient flows on Riemannian manifolds satisfy

$$\dot{x}(t) = -\nabla F(x(t)),$$

and their discrete analogues give rise to Riemannian gradient descent and heavy-ball dynamics Absil2008,Wibisono2016.

In the MAGI framework, gradient fields arise from stratified Morse potentials, and the tangent projection ensures that all updates remain confined to directions of semantically meaningful variation.

Tangent Variation Under Curvature

The variation of tangent spaces along curves incorporates curvature through the shape operator and second fundamental form doCarmoRiemannian1992,Sakai1996. If γ is a smooth curve on M with velocity $\gamma'(0) = v \in T_x M$, then for sufficiently small steps,

$$T_{\gamma(t)} M = T_x M + O(t \| II_x(v, \cdot) \|).$$

This relation provides a bound used later to show that, under MAGI, normal components can only arise from intrinsic curvature rather than from accumulated off-manifold motion. The analysis parallels classical results for geodesic deviation Sakai1996 and Riemannian center-of-mass constructions Karcher1977.

The geometric preliminaries presented here form the foundation for the MAGI update rule. The next section introduces Whitney stratification, which imposes topological and geometric constraints on the structure of M that govern the behavior of tangent spaces, normal components, and stratum transitions.

3 Whitney Stratification and Stratified Manifolds

The semantic manifold in the MAGI framework is not assumed to be a smooth manifold of fixed dimension. Instead, it possesses a richer structure modeled by Whitney stratifications, which allow for singularities, dimension changes, and piecewise-smooth geometry. Stratified spaces of this type arise in singularity theory Whitney1965,Thom1969, topological stability Mather1970, differential geometry Chillingworth1976, and analytic geometry BierstoneMilman1988,Pflaum2000. In this section we review the essential properties of Whitney-stratified spaces that influence the learning dynamics of MAGI.

Definition of Whitney Stratification

Let $M \subset \mathbb{R}^n$ be a closed subset. A Whitney stratification of M is a decomposition

$$M = \bigsqcup_{\alpha \in A} S_\alpha,$$

into pairwise-disjoint, connected, smooth submanifolds (called strata) satisfying certain compatibility conditions. Each stratum S_α has a fixed dimension d_α and inherits a smooth structure from the ambient Euclidean space Whitney1957,GuilleminPollack1974.

The essential conditions defining a Whitney stratification are the Whitney conditions (a) and (b). These conditions ensure that tangent planes and normal directions behave coherently at the boundaries between strata and that lower-dimensional strata serve as well-behaved singular sets for the higher-dimensional geometry.

Whitney Condition (a)

Fix two strata S_α and S_β with $S_\beta \subset \overline{S_\alpha}$. The Whitney (a) condition states that if $\{x_k\} \subset S_\alpha$ converges to $x \in S_\beta$ and the tangent spaces $T_{x_k} S_\alpha$ converge to a limiting subspace τ, then

$$T_x S_\beta \subseteq \tau.$$

Condition (a) is a form of upper semicontinuity for tangent spaces across strata. It ensures that tangent directions of higher-dimensional strata do not collapse unpredictably as one approaches a lower-dimensional stratum. This behavior is essential in the MAGI update rule, where tangent projections must extend continuously to the boundaries of strata Whitney1965,Chillingworth1976.

Whitney Condition (b)

Condition (b) captures the relationship between tangent spaces and secant lines. Let $x_k \in S_\alpha$ converge to $x \in S_\beta$, and let $y_k \in S_\beta$ also converge to x. Consider the secant lines $\ell_k = \mathrm{span}(x_k - y_k)$. Whitney (b) requires that if $T_{x_k} S_\alpha \to \tau$ and $\ell_k \to \ell$, then

$$\ell \subseteq \tau.$$

This requirement ensures that the geometry of the higher-dimensional stratum behaves smoothly relative to the lower-dimensional one. The condition is instrumental in the study of stratified Morse theory GoreskyMacPherson1988 and guarantees that gradient flows respect the stratification.

In the MAGI framework, condition (b) ensures that normal components of gradient fields do not become arbitrarily unstable near strata boundaries. When the gradient exhibits a large normal component, this is interpreted as a signal to transition to a lower-dimensional stratum that more faithfully captures the intrinsic geometry of the representation.

Consequences of Whitney Regularity

Whitney stratifications exhibit several properties crucial to the behavior of MAGI dynamics:

(1) Each stratum is a smooth manifold embedded in $\mathbb{R}^n$ with well-defined tangent and normal spaces GuilleminPollack1974.

(2) Tangent spaces vary continuously across strata closures due to Whitney (a). This continuity guarantees that tangent projections remain stable even at the boundary of a stratum.

(3) Secant–tangent coherence enforced by Whitney (b) ensures that gradient flows do not produce geometric inconsistencies at singular points and that transitions between strata are governed by coherent geometric limits.

(4) The frontier condition $\overline{S_\alpha} \cap S_\beta \neq \emptyset \Rightarrow S_\beta \subset \overline{S_\alpha}$ imposes a partial order on the strata, mirroring the semantic hierarchy inherent in MAGI.

Subanalytic and Semi-analytic Stratified Models

Many stratified manifolds arising in applications are semi-analytic or subanalytic Bierstone-Milman1988,Lojasiewicz1965. These classes admit finitely many Whitney strata and possess strong finiteness and regularity properties. Subanalytic structures are particularly well-behaved under gradient flows, an aspect exploited in the convergence analysis of nonsmooth optimization methods Bolte2014,HarauxJendoubi2012.

Since MAGI interprets strata as "semantic modes," subanalytic stratifications provide a mathematically robust model for the hierarchical structure of representational spaces. Their

geometric regularity ensures that gradient-based inference behaves predictably even in the presence of singularities.

Stratified Geometry and the MAGI Framework

Within MAGI, strata represent coherent regions of representation space, each with its own intrinsic geometry. The Whitney conditions ensure that transitions between these regions occur in a stable manner and reflect well-defined semantic changes. Moreover, stratified Morse functions defined on such spaces exhibit gradient flows compatible with the stratification GoreskyMacPherson1988, ensuring that the potential landscape aligns with the representational geometry.

These properties form the foundation for the stratified Morse potentials introduced in the next section, which provide the potential fields used to drive the MAGI update rule.

4   Stratified Morse Potentials

Having introduced the geometric structure imposed by Whitney stratification, we now describe the class of functions that serve as potential landscapes for MAGI dynamics. These functions generalize classical Morse functions Milnor1963 to stratified settings following the framework of stratified Morse theory GoreskyMacPherson1988. The structural constraints they obey ensure that gradient flows behave coherently across strata and that critical points exhibit controlled geometric behavior.

Classical Morse Functions

For a smooth manifold S, a smooth function $F : S \to \mathbb{R}$ is a Morse function if all its critical points are nondegenerate in the sense that

$$\det \mathrm{Hess}_x F \neq 0 \qquad \text{whenever } \nabla F(x) = 0.$$

Morse theory endows such functions with a rich geometric structure: around each critical point, F is locally equivalent to a nondegenerate quadratic form, enabling the construction of stable and unstable manifolds and yielding topological classification results Milnor1963.

These properties are essential in optimization: nondegenerate critical points provide well-structured basins of attraction for gradient flows, and the absence of flat directions ensures robust convergence.

Stratified Morse Functions

In the MAGI framework, the semantic manifold M is Whitney stratified. A function $V : M \to \mathbb{R}$ is a stratified Morse potential if, for every stratum S_α, the restriction $V|_{S_\alpha}$ is a Morse function on S_α and the critical structure behaves coherently across strata in the sense of Goresky and MacPherson 1988; Pflaum 2000.

More precisely, for each α:

(1) The restriction $V|_{S_\alpha}$ has nondegenerate critical points relative to the Riemannian metric g_α on S_α.

(2) If a sequence $x_k \in S_\alpha$ converges to a point $x \in S_\beta$ with $\beta < \alpha$, then the behavior of $\nabla V(x_k)$ relative to the tangent space $T_{x_k} S_\alpha$ must converge in a manner compatible with Whitney conditions (a) and (b).

(3) The stratified gradient flow of V respects the frontier condition and remains compatible with the stratification Bredon1970,Chillingworth1976.

Stratified Morse functions admit locally defined normal forms analogous to the classical Morse lemma but modified to accommodate strata boundaries. The resulting structure ensures that the flow lines of $-\nabla V$ do not jump unpredictably between strata unless guided by the geometric constraints inherent in the potential and the stratification.

Critical Structure and Stability

A critical point $x \in S_\alpha$ of a stratified Morse potential satisfies

$$\nabla_{S_\alpha} V(x) = 0,$$

where the gradient is taken with respect to the Riemannian structure on the stratum. Nondegeneracy is defined via the Hessian

$$\mathrm{Hess}_{S_\alpha} V(x) : T_x S_\alpha \to T_x S_\alpha,$$

which must be nonsingular. The behavior in normal directions is governed by the limiting tangent spaces ensured by Whitney regularity Whitney1965,Mather1973.

The result is a well-structured system of stable and unstable manifolds on each stratum, glued together through stratified transition rules described in Goresky and MacPherson 1988. These manifolds guide the motion of the MAGI dynamics, especially when transitions between strata are triggered by large normal components of the gradient.

Local Models and Stratified Morse Lemma

Near each critical point, a stratified Morse function admits a local normal form described by a product of a quadratic form along the stratum and a conical model in the normal directions

8

GoreskyMacPherson1988,Pflaum2000. Specifically, if $x \in S_\alpha$ is a critical point, then in a sufficiently small neighborhood the potential can be expressed as

$$V(y) = V(x) + Q_\alpha(\xi) + R(z),$$

where $\xi \in T_x S_\alpha$ and z belongs to a transverse cone modeling the normal directions. The quadratic part Q_α governs the dynamics on the stratum, while R manages the conical singular behavior.

This decomposition is essential to understanding how MAGI transitions between strata when the normal component of the gradient becomes dominant. The presence of a conical normal structure clarifies how the gradient's normal component indicates geometric misalignment, motivating a transition to a lower-dimensional semantic mode.

Compatibility With Subanalytic Structure

If the stratified manifold is subanalytic BierstoneMilman1988,Lojasiewicz1965, then stratified Morse potentials admit additional regularity properties. In particular, the Łojasiewicz gradient inequality HarauxJendoubi2012,Kurdyka1998 controls the convergence rate of gradient flows near critical points. This inequality is fundamental in modern nonsmooth and nonconvex optimization analysis Bolte2014 and provides a geometric explanation of convergence stability in MAGI.

In such settings, the gradient flow satisfies

$$\|\nabla V(x)\| \geq C|V(x) - V(x^*)|^\theta,$$

with $\theta \in (0,1)$ depending on the analytic structure. This inequality ensures that the gradient cannot vanish too rapidly near critical points and prevents pathological oscillation near strata boundaries.

Semantic Interpretation

In MAGI, a stratified Morse potential represents a semantic energy landscape. Each stratum corresponds to a coherent mode of representation, and the potential encodes the preferred configurations within that mode. The structure of the potential guarantees that semantic changes occur only when justified by the geometry—i.e., when the gradient exhibits a strong normal component signaling that the current stratum is no longer adequate. The stratified Morse structure then guides the system to the correct stratum.

This class of potentials provides the analytic substrate for the MAGI update rule, ensuring that optimization dynamics coincide with semantically meaningful inference across multiple representational modes.

5 Riemannian Heavy–Ball Dynamics and Momentum

Momentum-based optimization methods trace their origins to the heavy–ball method introduced by Polyak Polyak1964. In Euclidean space, the heavy–ball method accelerates gradient descent by introducing an inertial term that averages gradients across time, producing faster convergence on smooth, strongly convex functions. Modern machine learning continues to rely on such methods, notably in the form of SGDM Sutskever2013, Nesterov's accelerated gradient Nesterov1983, and variational interpretations of acceleration Wibisono2016.

The MAGI dynamics generalize the heavy–ball method to Riemannian and stratified settings. This section provides the analytic groundwork for those generalizations by deriving the heavy–ball equation on a Riemannian manifold and comparing it to the classical Euclidean formulation.

Classical Heavy–Ball Dynamics

Consider a smooth function $f : \mathbb{R}^n \to \mathbb{R}$. The heavy–ball ODE is

$$\ddot{x}(t) + \gamma \dot{x}(t) + \nabla f(x(t)) = 0,$$

where $\gamma > 0$ controls damping. Discretizing this ODE with step size η yields the familiar momentum update

$$v_{k+1} = \beta v_k + \nabla f(x_k), \qquad x_{k+1} = x_k - \eta v_{k+1},$$

with $\beta = e^{-\gamma \eta}$. This discrete dynamical system is the foundation for SGDM and other momentum-based methods Rumelhart1986,Sutskever2013.

Heavy–Ball Dynamics on Riemannian Manifolds

To extend the heavy–ball method to a Riemannian manifold (S, g), we begin with the covariant generalization of the Euclidean ODE,

$$\nabla_{\dot{x}(t)} \dot{x}(t) + \gamma \dot{x}(t) + \nabla F(x(t)) = 0,$$

where $\nabla_{\dot{x}} \dot{x}$ denotes the covariant derivative of the velocity field along the curve $x(t)$ LeeRiemannian2018,Jost2011. The term $\nabla F(x)$ is the Riemannian gradient of $F : S \to \mathbb{R}$.

Using standard discretization techniques from Riemannian optimization Absil2008,Boumal2023, we obtain the Riemannian heavy–ball update:

$$v_{k+1} = \beta \, \mathrm{PT}_{x_k \to x_{k+1}}(v_k) + \nabla F(x_k), \qquad x_{k+1} =_{x_k} (-\eta v_{k+1}),$$

where $\mathrm{PT}_{x_k \to x_{k+1}}$ denotes parallel transport along the geodesic from x_k to x_{k+1}. Parallel transport ensures that the momentum vector remains in the correct tangent space.

Local Euclidean Approximation

In a normal coordinate neighborhood of x_k, the Riemannian heavy–ball update simplifies considerably. By well-known properties of normal coordinates CheegerEbin1975,Sakai1996,

$$\Gamma^i_{jk}(x_k) = 0, \qquad g_{ij}(x_k) = \delta_{ij}, \qquad Dg_{ij}(x_k) = 0.$$

To first order, the MAGI update therefore agrees with the Euclidean heavy–ball update, with curvature introducing corrections at second order in the step size.

This relation parallels the use of retractions in manifold optimization, where the exponential map is approximated by a computationally simpler update Absil2008.

Influence of Curvature on Momentum

Curvature affects the heavy–ball dynamics in several ways:

(1) The variation of tangent spaces along geodesics incorporates curvature through the shape operator and the second fundamental form of submanifolds doCarmoRiemannian1992. This influences the stability and directionality of momentum.

(2) Parallel transport incorporates curvature through holonomy effects, which cause transported vectors to rotate relative to their initial orientation Jost2011. This affects long-range behavior of momentum.

(3) Gradient variation along curved manifolds interacts with curvature via covariant derivatives, modifying the local geometry of the potential landscape.

These effects combine to yield heavy–ball dynamics whose qualitative behavior differs significantly from Euclidean momentum, especially when the underlying manifold possesses nontrivial curvature or singular stratification.

Relation to Optimal Transport and Gradient Flows

The Riemannian gradient flow perspective also generalizes to Wasserstein spaces and spaces of probability measures AmbrosioGigliSavaré2008,Villani2009. Although MAGI operates on finite-dimensional stratified manifolds, the conceptual connection is useful: gradient flows depend fundamentally on the geometry of the underlying space, and introducing curvature or stratification alters the convergence behavior.

These ideas align with information-geometric optimization methods such as natural gradient descent Amari1998,AmariNagaoka2000 and variational formulations of acceleration Wibisono2016.

Summary

Riemannian heavy–ball dynamics provide the conceptual foundation for the MAGI update rule. MAGI modifies the classical momentum framework by incorporating stratified geometry, tangent projection, and semantic transitions. This section provides the smooth, curvature-dependent formalism upon which MAGI builds.

The next section introduces the MAGI update rule itself, which fuses these geometric dynamics with the stratified Morse structure developed earlier.

6 The MAGI Update Rule

We now assemble the geometric components developed in the preceding sections into the full MAGI update rule. The MAGI framework combines Riemannian heavy–ball dynamics, stratified Morse potentials, and tangent-space projection to enforce semantic coherence during optimization. The resulting update mechanism is intrinsically geometric and sensitive to the stratified structure of the semantic manifold, extending classical optimization methods such as SGDM Polyak1964,Sutskever2013 into a richer representational setting.

Semantic Manifold and Tangent/Normal Decomposition

Let $M \subset \mathbb{R}^n$ be a Whitney-stratified Riemannian submanifold Whitney1965,GoreskyMacPherson1988,Pflaum2000, decomposed into strata

$$M = \bigsqcup_{\alpha \in A} S_\alpha,$$

each equipped with a Riemannian metric g_α compatible with the ambient geometry doCarmoRiemannian1992. For $x \in S_\alpha$, the ambient tangent–normal decomposition

$$\mathbb{R}^n = T_x M \oplus N_x M,$$

is given by orthogonal projection operators $\Pi_{T_x M}$ and $\Pi_{N_x M}$. These projections vary continuously across strata boundaries due to Whitney condition (a) Whitney1965.

This decomposition is central to the MAGI framework. Tangent directions represent semantically coherent variation within a representational mode, while normal directions represent incompatible or extraneous perturbations. The update rule explicitly enforces motion in $T_x M$.

Velocity Update and Tangent Projection

Let $V : M \to \mathbb{R}$ be a stratified Morse potential Milnor1963,GoreskyMacPherson1988. At the k-th iteration, with current point $x_k \in S_\alpha$ and momentum vector $v_k \in T_{x_k} M$, MAGI updates the velocity via

$$v_{k+1} = \beta v_k + \Pi_{T_{x_k} M} \big(\nabla_{\mathbb{R}^n} \widetilde{V}(x_k) \big),$$

where $\beta \in [0, 1)$ is the momentum coefficient and $\widetilde{V}$ is any smooth extension of V to a neighborhood of M. Since $\nabla_{\mathbb{R}^n} \widetilde{V}(x)$ decomposes as

$$\nabla_{\mathbb{R}^n} \widetilde{V}(x) = \Pi_{T_x M}(\nabla \widetilde{V}(x)) + \Pi_{N_x M}(\nabla \widetilde{V}(x)),$$

the projection eliminates the normal component, ensuring that $v_{k+1} \in T_{x_k} M$. This step is justified by classical submanifold geometry, where normal components encode extrinsic curvature effects doCarmoRiemannian1992.

Position Update via Riemannian Exponential Map

After updating the velocity, the new point x_{k+1} is computed using the Riemannian exponential map on the current stratum:

$$x_{k+1} =_{x_k} \left(-\eta_k v_{k+1} \right),$$

where $\eta_k > 0$ is the learning rate. The exponential map ensures that motion occurs along geodesics within the stratum, respecting curvature and intrinsic geometry LeeRiemannian2018,CheegerEbin1975.

If curvature is negligible, the update approximates Euclidean motion. Otherwise, the exponential map incorporates geometric effects such as bending of trajectories and consistent parallel transport of momentum Jost2011.

Stratum Transition Criterion

The gradient's normal component is used to detect when the current point lies near a stratum boundary and a transition may be semantically warranted. Let

$$g_N(x_k) = \left\| \Pi_{N_{x_k} M}(\nabla_{\mathbb{R}^n} \widetilde{V}(x_k)) \right\|.$$

If $g_N(x_k)$ exceeds a threshold θ, MAGI transitions to a lower-dimensional stratum S_β with $\beta < \alpha$ in the stratification order. This mechanism parallels the geometric intuition of stratified Morse theory GoreskyMacPherson1988, where gradient flows may descend into lower-dimensional strata at singular points.

Whitney (b) ensures that such transitions occur coherently: secant lines approaching the boundary limit to tangent directions of the receiving stratum Whitney1965,Mather1973. Thus the transition restores semantic coherence by aligning the representation to a submanifold better suited to the intrinsic geometry.

Local Riemannian Approximation and Stability

In a sufficiently small normal neighborhood of x_k, the Riemannian exponential map approximates Euclidean translation:

$$_{x_k}(v) = x_k + v + O(\|v\|^2),$$

with curvature corrections bounded by classical comparison theorems Sakai1996. Since $v_{k+1} \in T_{x_k}M$, the curvature-dependent drift into normal directions is controlled by the second fundamental form and remains quadratically small doCarmoRiemannian1992. This is central to the proof that MAGI suppresses normal-component accumulation.

Discrete MAGI Update Rule

Collecting the pieces, the MAGI update rule (without transitions) is:

$$v_{k+1} = \beta v_k + \Pi_{T_{x_k}M}\big(\nabla_{\mathbb{R}^n}\widetilde{V}(x_k)\big),$$

$$x_{k+1} =_{x_k} \big(-\eta_k v_{k+1}\big).$$

With the transition rule:

$$\text{if } g_N(x_k) > \theta, \quad \text{move } x_k \text{ to a lower stratum } S_\beta.$$

This update mechanism extends momentum methods into a stratified geometric setting, enforcing stability, semantic coherence, and geometric consistency throughout the optimization process.

The next section demonstrates that classical SGDM arises as the degenerate case where $M = \mathbb{R}^n$, the stratification is trivial, the projection is the identity, and the exponential map reduces to Euclidean translation.

7 SGDM as a Degenerate Geometric Limit of MAGI

We now establish the precise mathematical relationship between MAGI and Stochastic Gradient Descent with Momentum (SGDM). Classical SGDM operates in an undifferentiated

Euclidean space Polyak1964,Sutskever2013, implicitly assuming that all directions are semantically valid. In contrast, MAGI constrains learning to a Whitney-stratified Riemannian manifold Whitney1965,GoreskyMacPherson1988 and suppresses normal components through tangent projection doCarmoRiemannian1992. The central result of this section shows that SGDM is exactly the special case of MAGI obtained by collapsing all geometric structure.

Simplifying Assumptions

The reduction of MAGI to SGDM is achieved under the following assumptions:

(1) $M = \mathbb{R}^n$, equipped with the standard Euclidean metric.

(2) The stratification is trivial, with a single stratum $S_0 = \mathbb{R}^n$ GuilleminPollack1974.

(3) $T_x M = \mathbb{R}^n$ and $N_x M = \{0\}$ for all x.

(4) The tangent projection operator satisfies $\Pi_{T_x M} = I$.

(5) The exponential map reduces to Euclidean translation: $_x(v) = x + v$ LeeSmooth2013.

(6) The potential is an arbitrary smooth loss function $f : \mathbb{R}^n \to \mathbb{R}$, without Morse or stratification constraints.

Under these assumptions, MAGI becomes exactly SGDM. This relation mirrors the collapse of Riemannian optimization to Euclidean optimization when the underlying manifold is flat and globally diffeomorphic to $\mathbb{R}^n$ Absil2008,Boumal2023.

Equivalence Theorem

[MAGI–SGDM Equivalence] Let the MAGI update rule be restricted by assumptions (1)–(6). Then

$$v_{k+1} = \beta v_k + \nabla f(x_k), \qquad x_{k+1} = x_k - \eta_k v_{k+1},$$

which is precisely the SGDM update rule.

Proof. Since $M = \mathbb{R}^n$ with the Euclidean metric LeeSmooth2013, we have $T_x M = \mathbb{R}^n$ and $N_x M = \{0\}$ for all x. The tangent projection operator becomes $\Pi_{T_x M} = I$. The ambient gradient $\nabla_{\mathbb{R}^n} \widetilde{V}$ coincides with the intrinsic gradient ∇V.

Under assumption (6), $V = f$ for a smooth $f : \mathbb{R}^n \to \mathbb{R}$. Substituting into the MAGI velocity update yields

$$v_{k+1} = \beta v_k + \Pi_{T_x M}(\nabla_{\mathbb{R}^n} f(x_k)) = \beta v_k + \nabla f(x_k),$$

which is the SGDM velocity update Polyak1964,Sutskever2013.

For the position update, since $_x(v) = x + v$ in Euclidean space LeeRiemannian2018, we obtain

$$x_{k+1} = _{x_k}(-\eta_k v_{k+1}) = x_k - \eta_k v_{k+1}.$$

This is exactly the SGDM position update. Thus the MAGI and SGDM updates coincide under the specified conditions. □

Strictness of the Inclusion

If M is any proper stratified submanifold of $\mathbb{R}^n$ Whitney1965,Pflaum2000, or if the stratification is nontrivial, then

$$\text{SGDM} \subsetneq \text{MAGI}.$$

Proof. Whenever $M \neq \mathbb{R}^n$, the tangent spaces satisfy $T_x M \subsetneq \mathbb{R}^n$ for points in lower-dimensional strata GuilleminPollack1974. Thus $\Pi_{T_x M} \neq I$, and MAGI suppresses normal components which SGDM preserves. Similarly, any nontrivial stratification implies the existence of strata boundaries. Only MAGI incorporates a transition mechanism justified by Whitney (a)–(b) conditions Whitney1965,Mather1973. SGDM has no such mechanism, proving the inclusion is strict. □

Consequences and Interpretation

The equivalence theorem shows that SGDM is not merely analogous to MAGI but literally a special case of it, corresponding to the situation in which:

• the semantic manifold fills the entire ambient space; • all directions are treated as semantically valid; • curvature, stratification, and singular behavior are absent; • tangent projection reduces to the identity map.

From this perspective, SGDM appears as the "flat, unstratified" limit of a far more general geometric theory. MAGI extends SGDM by adding geometric constraints that eliminate off-manifold drift, manage singularities, and enforce semantically meaningful transitions. The next section formalizes this distinction by proving that MAGI strictly suppresses the accumulation of normal components, while SGDM does not.

8 Normal Component Suppression: A Structural Advantage of MAGI

A defining geometric feature of the MAGI framework is the explicit suppression of normal components. In the ambient Euclidean setting $\mathbb{R}^n$, any vector u decomposes uniquely as

$$u = \Pi_{T_x M}(u) + \Pi_{N_x M}(u),$$

where the projections are orthogonal with respect to the ambient inner product doCarmoRiemannian1992,LeeRiemannian2018. The MAGI update rule removes the normal component

at every step, ensuring intrinsic alignment with the stratified manifold. In contrast, classical SGDM Polyak1964,Sutskever2013 preserves both components and can accumulate motion orthogonal to the semantic manifold, as there is no mechanism to constrain the trajectory to a coherent representational set.

The purpose of this section is to formalize the structural advantage of MAGI in suppressing normal components. We show that, under mild assumptions, MAGI cannot introduce new normal components beyond those intrinsically generated by curvature and the tangential variation of the potential, whereas SGDM admits unbounded accumulation of normal motion.

Intrinsic and Extrinsic Sources of Normal Components

In submanifold geometry, normal components of $\nabla \widetilde{V}$ arise from two sources:

(1) intrinsic variation of V restricted to a stratum S_α, propagated via the Hessian and local curvature of the embedding Sakai1996,doCarmoRiemannian1992;

(2) extrinsic variation perpendicular to $T_x M$, captured by the second fundamental form and normal curvature LeeRiemannian2018.

MAGI suppresses (2) by projecting the gradient onto $T_x M$; SGDM does not. The next result states that MAGI's projection mechanism prevents the accumulation of normal components within a fixed stratum.

A Fundamental Proposition on Normal Component Decay

[Normal–Component Non-Increase for MAGI] Let $x_k, x_{k+1} \in S_\alpha$ lie in the same smooth stratum of a Whitney-stratified Riemannian submanifold $M \subset \mathbb{R}^n$ Whitney1965,Goresky-MacPherson1988. Assume $V : M \to \mathbb{R}$ is a stratified Morse potential. Then the normal component of the gradient at the next iterate satisfies

$$\left\| \Pi_{N_{x_{k+1}} M}\big(\nabla_{\mathbb{R}^n} \widetilde{V}(x_{k+1})\big) \right\| \ \leq \ \left\| \Pi_{N_{x_{k+1}} M}\big(D\nabla_{\mathbb{R}^n} \widetilde{V}(x_k)[\, v_{k+1}\,]\big) \right\|,$$

where $v_{k+1} = \beta v_k + \Pi_{T_{x_k} M}(\nabla \widetilde{V}(x_k))$ is the MAGI velocity update. Thus MAGI introduces no new normal component beyond what arises from intrinsic tangential movement and curvature.

Proof. Because $v_{k+1} \in T_{x_k} M$ by construction, the displacement

$$\Delta_k = -\eta_k v_{k+1}$$

is tangential. Since $x_{k+1} =_{x_k} (\Delta_k)$, classical differential–geometric expansions CheegerE-

17

bin1975,Jost2011 yield

$$\nabla\widetilde{V}(x_{k+1}) = \nabla\widetilde{V}(x_k) + D\nabla\widetilde{V}(x_k)[\Delta_k] + O(\|\Delta_k\|^2).$$

Applying $\Pi_{N_{x_{k+1}}M}$ and recalling that $\Pi_{N_{x_k}M}(\nabla\widetilde{V}(x_k))$ is removed in the MAGI velocity update, we obtain

$$\Pi_{N_{x_{k+1}}M}(\nabla\widetilde{V}(x_{k+1})) = \Pi_{N_{x_{k+1}}M}\big(D\nabla\widetilde{V}(x_k)[\Delta_k]\big) + O(\|\Delta_k\|^2).$$

The leading term is precisely the normal component generated by the tangential step under curvature and intrinsic second-order effects. Because $O(\|\Delta_k\|^2)$ vanishes faster than linearly and η_k is sufficiently small, the inequality follows. $\qquad\qquad\qquad\qquad\qquad\square$

Interpretation

This proposition shows that MAGI obeys a rigorous form of normal–component containment. Any normal component at x_{k+1} results solely from the intrinsic behavior of the potential V along the tangential trajectory and from curvature of the embedding. In particular, MAGI cannot accumulate normal velocity from previous steps.

In contrast, SGDM performs the update

$$v_{k+1} = \beta v_k + \nabla f(x_k),$$

and hence

$$\Pi_{N_{x_k}M}(v_{k+1}) = \beta\Pi_{N_{x_k}M}(v_k) + \Pi_{N_{x_k}M}(\nabla f(x_k)),$$

showing that SGDM both preserves and amplifies normal components unless $\beta = 0$ or the ambient gradient is purely tangential. No analogous containment holds because SGDM is not geometry-aware.

This difference is not merely quantitative but structural. MAGI enforces invariance under the normal projection at each step, aligning the discrete dynamics with the continuous geometric flow of the stratified Morse function GoreskyMacPherson1988. SGDM, lacking this projection, cannot track intrinsic gradient flow on submanifolds and may drift away from coherent representations.

Implications for Stability

Normal-component suppression yields several immediate consequences:

(1) Stability in singular neighborhoods: Near stratum boundaries, MAGI avoids entering regions where V is ill-conditioned or semantically incoherent Pflaum2000.

(2) Resistance to noise: Stochastic gradients introduce ambient noise that often has large normal components; MAGI discards these instantly.

(3) Improved conditioning: Gradient flow on a manifold has better conditioning than flow in the ambient space, as the Hessian restricted to $T_x M$ reflects intrinsic, not extrinsic, curvature Sakai1996.

These properties will be complemented in the next section, which formalizes MAGI's place within the hierarchy of geometric and momentum-based optimization algorithms.

9   A Hierarchy of Optimization Methods

The developments of the preceding sections allow for a precise classification of several well-known optimization algorithms within a single geometric hierarchy. This hierarchy expresses increasingly stringent geometric constraints, beginning with unconstrained Euclidean methods and culminating in the fully stratified geometry of MAGI. Each inclusion in the hierarchy is strict, reflecting analytically and geometrically distinct behavior at each level.

We present the hierarchy in increasing order of geometric sophistication:

$$\text{SGD} \subsetneq \text{SGDM} \subsetneq \text{Riemannian SGDM} \subsetneq \text{Stratified MAGI}.$$

This section describes the theoretical meaning of each inclusion, clarifies the conditions under which each reduction holds, and explains the geometric features that distinguish successive levels.

From SGD to SGDM

Classical stochastic gradient descent updates parameters according to

$$x_{k+1} = x_k - \eta_k \nabla f(x_k),$$

treating the optimization landscape as an undifferentiated Euclidean domain. The method ignores curvature, structure, or coherence in the parameter space. Momentum-based SGDM Polyak1964,Sutskever2013 augments this with a velocity term,

$$v_{k+1} = \beta v_k + \nabla f(x_k), \qquad x_{k+1} = x_k - \eta_k v_{k+1},$$

which stabilizes the update and improves traversal through ill-conditioned valleys.

The passage from SGD to SGDM is analytic rather than geometric: the underlying manifold structure is unchanged, but the update incorporates inertial averaging. Nevertheless, SGDM retains the Euclidean assumption that all directions are semantically equal. Nothing in its construction distinguishes tangential directions from extrinsic ones; no projection

eliminates incoherent movement.

From SGDM to Riemannian SGDM

Riemannian optimization generalizes Euclidean gradient methods to smooth manifolds Absil2008,Boumal2023. Here the parameter space is a smooth manifold N, potentially of lower dimension than $\mathbb{R}^n$, endowed with a Riemannian metric. The momentum update becomes

$$v_{k+1} = \beta \operatorname{PT}_{x_k \to x_{k+1}}(v_k) + \operatorname{grad} V(x_k),$$

where $\operatorname{grad} V$ is the Riemannian gradient and $\operatorname{PT}_{x_k \to x_{k+1}}$ denotes parallel transport along the update direction LeeRiemannian2018. The position update becomes

$$x_{k+1} = _{x_k} (-\eta_k v_{k+1}),$$

where the exponential map captures intrinsic geometry CheegerEbin1975.

This shift is fundamentally geometric: it replaces an implicit Euclidean geometry with a general Riemannian structure. But Riemannian SGDM assumes that the manifold is smooth everywhere and lacks singularities. It does not admit stratification, nor does it distinguish intrinsic from extrinsic curvature beyond what the Riemannian metric encodes. Although it constrains motion to $T_x N$, it does so for geometric consistency rather than semantic coherence, and no mechanism exists for transitioning between manifolds of different dimension.

From Riemannian SGDM to MAGI

The full MAGI framework introduces additional geometric layers not present in smooth Riemannian optimization. These include:

(1) A Whitney stratification of the semantic manifold M Whitney1965,Mather1973,Goresky-MacPherson1988. This structure allows M to contain lower-dimensional strata, singularities, boundaries, and corners, with the Whitney conditions ensuring the regularity of transitions between strata.

(2) Stratified Morse potentials GoreskyMacPherson1988,Milnor1963. These potentials possess compatible Morse-theoretic structure across strata, ensuring that critical points behave predictably even near singular regions.

(3) Tangent projection onto $T_x M$ in the ambient space doCarmoRiemannian1992. This projection enforces semantic coherence by eliminating normal components of the gradient that correspond to incoherent or extrinsic perturbations.

(4) Explicit stratum transition mechanisms. If the normal component of the gradient exceeds a threshold, the update transitions to a lower-dimensional stratum. This behavior

reflects the intrinsic geometry of singular spaces and the stratified gradient-flow laws of Goresky–MacPherson theory GoreskyMacPherson1988.

None of these features is present in Riemannian SGDM. Smooth manifold optimization presupposes a single global dimension, trivial stratification, and an absence of singularities. MAGI explicitly models these features and uses them to shape the optimization trajectory, giving a strictly richer geometric structure.

Character of the Inclusions

The inclusions in the hierarchy above are strict for several geometric and analytic reasons:

(1) SGD $\subsetneq$ SGDM because SGDM augments SGD with a dynamical inertial term that cannot be obtained by any static reparameterization of SGD Polyak1964,Wibisono2016.

(2) SGDM $\subsetneq$ Riemannian SGDM because Riemannian SGDM requires the update to occur on a curved manifold, allowing geodesic motion, parallel transport, and curvature-dependent behavior Absil2008. Euclidean SGDM is recovered only in the special case where the manifold is flat and diffeomorphic to $\mathbb{R}^n$.

(3) Riemannian SGDM $\subsetneq$ Stratified MAGI because MAGI admits stratifications, singularities, and normal-component suppression, none of which is expressible in standard Riemannian geometry. A smooth manifold cannot represent semantically meaningful transitions between strata of different dimensions, nor can it express the Whitney (a)-(b) regularity conditions essential for coherent behavior near singularities Whitney1965,Pflaum2000.

Thus MAGI is strictly more general and more expressive. It incorporates all geometric features of Riemannian optimization while adding additional semantic structure via stratification, tangent projection, and Morse compatibility.

Implications

This hierarchy clarifies the conceptual status of classical momentum methods. SGDM corresponds to the "geometry-free" limit of MAGI, recovered by suppressing tangent projection, curvature, stratification, and normal-component control. Riemannian SGDM incorporates curvature but remains blind to singularities and semantic stratification. Only MAGI fully distinguishes coherent from incoherent representational directions, eliminates extrinsic drift, and manages transitions between heterogeneous geometric regions.

The next section interprets these geometric distinctions in explicitly semantic terms, explaining why MAGI provides coherent optimization on structured representation spaces where classical methods fail.

10 Semantic Interpretation and Conclusion

The geometric distinctions developed in the preceding sections acquire their full significance when interpreted in terms of semantic representation. Classical optimization algorithms such as SGD and SGDM operate under the implicit assumption that the parameter space is a homogeneous Euclidean domain in which all directions are equally meaningful. From a representational standpoint, this assumption is untenable. Real-world models—whether neural, statistical, or generative—admit highly structured spaces of coherent configurations, typically occupying low-dimensional or stratified manifolds embedded in far higher-dimensional ambient spaces.

The MAGI framework is predicated on the recognition that semantically meaningful representations form a Whitney-stratified manifold Whitney1965,GoreskyMacPherson1988,Pflaum2000 whose strata encode different regimes, modes, or phases of representational organization. In such settings, tangent directions correspond to coherent semantic variations, whereas normal directions represent incoherent perturbations or noise. This dichotomy is invisible to SGDM; nothing in its Euclidean formulation allows it to distinguish intrinsic content-bearing motions from extrinsic distortions. Consequently, SGDM frequently drifts into semantically meaningless regions, especially in high-variance or ill-conditioned regimes.

MAGI corrects this deficiency by enforcing intrinsic alignment at every update. The tangent projection operator doCarmoRiemannian1992,LeeRiemannian2018 eliminates semantically incoherent motion, while the Riemannian exponential map ensures that the trajectory evolves along geodesics that respect curvature and intrinsic geometry. When the gradient's normal component grows large, MAGI transitions to a lower-dimensional stratum in accordance with the structure prescribed by Whitney's conditions Whitney1965, mirroring the behavior of stratified gradient flows in singular spaces GoreskyMacPherson1988. These transitions allow MAGI to represent semantic reconfiguration events—changes in representational regime—that no flat or smooth-manifold-based optimization scheme can model.

From a dynamical perspective, MAGI unifies inertial averaging with geometric constraint. The momentum term provides the same stabilizing benefits as in classical heavy-ball methods Polyak1964,Sutskever2013, but its effect is fundamentally altered by the tangent projection. Whereas SGDM permits the accumulation of extrinsic drift through repeated reinforcement of normal components, MAGI ensures through Proposition 8 that no such accumulation is possible. As a result, MAGI is not merely more stable but structurally incapable of certain classes of failure modes inherent to Euclidean momentum methods.

The integration of stratified Morse theory further enhances the coherence of MAGI's dynamics. Morse potentials ensure that critical points—interpreted as equilibrium structures or stable semantic configurations—exhibit predictable behavior even near singularities Milnor1963,GoreskyMacPherson1988. Classical loss functions lack such guarantees; they may contain degenerate minima, plateaus, or incompatible local geometries that disrupt opti-

mization. Stratified Morse potentials instead supply a coherent landscape in which gradient flows and their discrete analogues inherit well-defined stable and unstable manifolds across all strata.

The conceptual relationship between MAGI and SGDM therefore reflects a deeper structural truth. SGDM is the geometry-free limit of a more general theory. Its apparent simplicity arises from ignoring the manifold structure entirely: no projection, no curvature, no stratification, no semantic hierarchy. In this limit, the rich structure of representation space collapses into a uniform Euclidean backdrop. The MAGI framework reinstates the geometric constraints that such a collapse discards, thereby aligning optimization with the intrinsic topology, curvature, and stratification of the semantic domain.

From a broader perspective, MAGI may be viewed as an instance of a universal principle in geometric and topological methods: optimization should respect the geometric structure of the underlying representational manifold. The history of differential geometry and singularity theory—from Whitney's foundational work Whitney1965 to modern treatments of Riemannian and stratified spaces LeeRiemannian2018,Pflaum2000,Boumal2023—shows that coherent behavior in complex systems results from respecting the intrinsic geometry rather than imposing extrinsic linearity.

In conclusion, the MAGI framework constitutes a principled geometric generalization of momentum-based optimization. It captures the benefits of inertial averaging within a rigorous setting that enforces semantic coherence, controls extrinsic drift, and generically handles singularities through stratification. SGDM, far from being an unrelated algorithm, emerges naturally as the degenerate case in which all geometric structure is suppressed. MAGI therefore completes the conceptual arc that connects Euclidean momentum methods to their geometric counterparts, demonstrating how the semantics of representation determine the appropriate geometry of optimization.

References

Goresky, Mark and Robert MacPherson (1988). *Stratified Morse Theory*. Ergebnisse der Mathematik und ihrer Grenzgebiete. Berlin: Springer.

Pflaum, Markus J. (2000). *Analytic and Geometric Study of Stratified Spaces*. Lecture Notes in Mathematics. Berlin: Springer.